How near are urban inhabitants to appreciated natural areas? An exploration of Hotspotmonitor based well-being indicators. Results for the Netherlands, Germany, and Denmark.

Sijtsma, Frans; Daams, Michiel

How near are urban inhabitants to appreciated natural areas?

An exploration of Hotspotmonitor based well-being indicators

Results for the Netherlands, Germany, and Denmark

Informal Report for OECD | Groningen, June 11, 2014

Frans J. Sijtsma and Michiel N. Daams (f.j.sijtsma@rug.nl / m.n.daams@rug.nl)

1 - Introduction

The OECD's project How's Life? paints a comprehensive picture of well-being through looking at people's material living conditions and quality of life across the population.[1] To assess with more (spatial) precision the regional differentiation in quality of life the OECD initiated the project How's life in your region?[2] At a Groningen workshop of this latter project Frans Sijtsma presented results from research using the Hotspotmonitor; an online survey tool which measures appreciation for natural areas. As a follow-up of the workshop it was decided to explore the possible use of Hotspotmonitor data to measure the 'perceived amenity value' of the natural areas within the reach of inhabitants of urban areas.

This report describes an exploration of indicators which build on the perceived amenity value of natural space to indicate regional differences in urban well-being. The subjective (yet systematic) character of the Hotspotmonitor based indicators deviates from objective indicators which are based on natural land use data, which lead to indicators which assume that well-being is constant across the type of nature analyzed. Hotspotmonitor based indicators may be used as complementary well-being indicators next to objective indicators as these may show different amenity values of nature as subjective valuations of amenity values are taken into account.

As an indicator for nature related well-being we focus on both the general presence of appreciated natural areas in FUAs and the mean distance of each FUAs population's to collectively appreciated natural areas. To link up to OECDs work in redefining urban areas this exploration uses the FUA, a functional urban areas definition of agglomerations (>50.000)[3] and their hinterland. FUAs[4] across the Netherlands, Germany, and Denmark are analyzed.

The remainder of the report is organized as follows. In Chapter 2 the data are discussed briefly and a methodological summary is given. Indicators are discussed one by one in Chapter 3, followed by a listing of discussion points in Chapter 4. Please note that a Technical appendix is included, which gives detailed information on the processing of data and the calculation of the indicators presented.

---

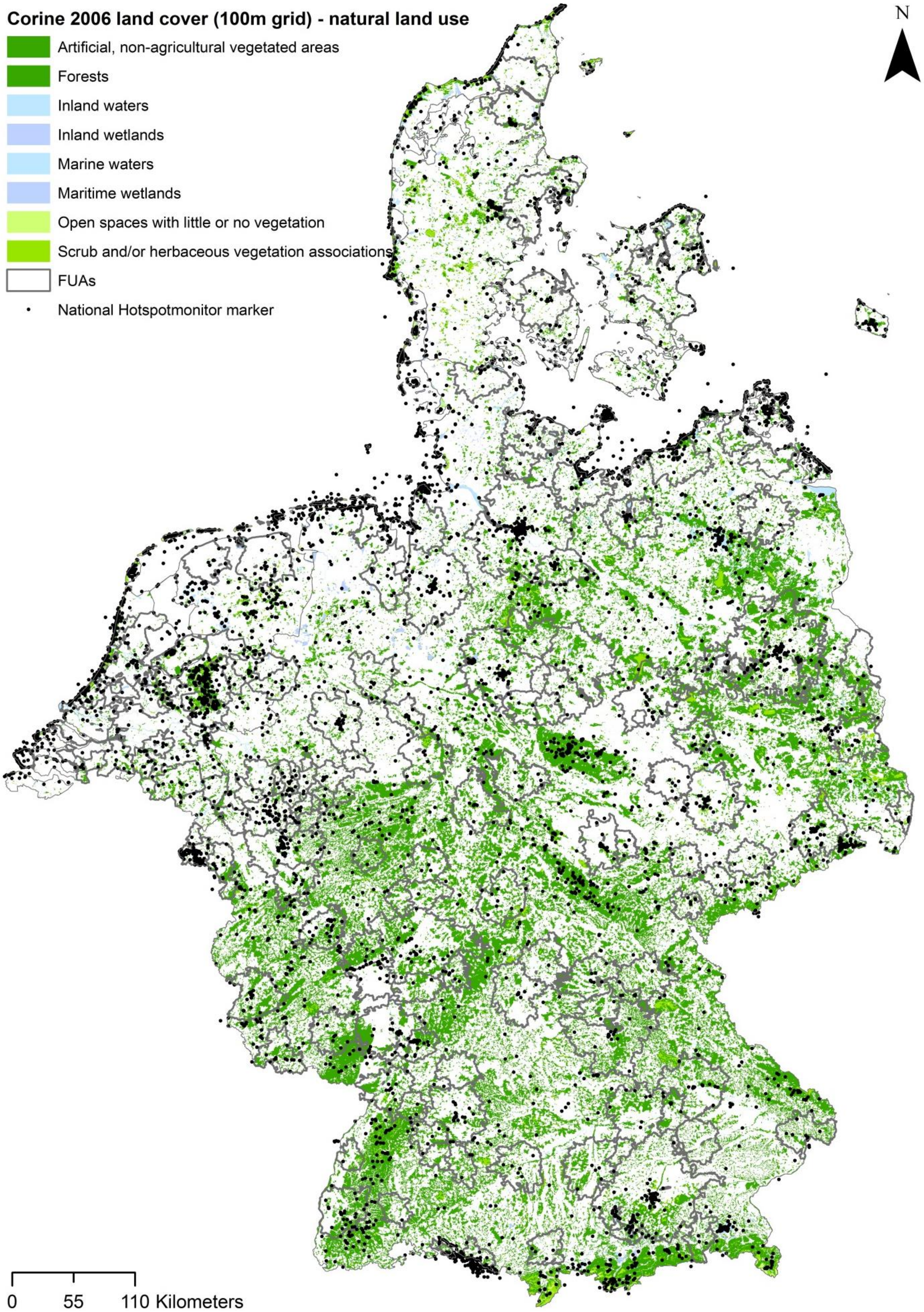
[1] http://www.oecd.org/statistics/howslife.htm

[2] For a full description of the aims see: http://www.oecd.org/regional/how-is-life-in-your-region.htm

[3] This figure applies to European agglomerations specifically.

[4] Core and hinterland are merged for each FUA for a more straightforward spatial interpretation of indicator values.

A view on the raw data: natural areas and perceived amenity value (markers)



2 – Data and methodological summary

2.1 List of datasets

- FUAs (N=148); polygon shapefile
    - Source: OECD

- HSM GfK 2013 (N=7,169); point shapefile
    - Source: Hotspotmonitor database June 2014

- Population in 2006 on a 1x1km grid (*the centroids are used*); polygon shapefile
    - Source: OECD

- Corine 2006 land use 100x100m grid; raster
    - Source: European Environment Agency

All data are (re)projected in the ETRS89_LAEA coordinate system.

2.2 Hotspotmonitor data

The Hotspotmonitor (HSM) survey measures each respondent’s perception of natural spaces amenity value on a local scale (2km from the respondent’s home), regional scale (20km from home), national, and international scale. For each scale separately, HSM-survey respondents are asked to mark the single natural space they perceive as highly valuable. The survey output includes point-location XY-coordinates of the markers which respondents have placed to pinpoint natural areas (on both land and water), as well as the XY-coordinates of their (approximate) living location. Since at the national scale HSM respondents can mark all natural spaces consistently we work with these markers only in this report’s exploration of well-being indicators at FUA-level. Such indicators could also be constructed using local and regional HSM markers, however this requires sufficient spatial coverage of HSM respondents across the FUAs analyzed.

Our working dataset of HSM markers includes 1,264 Dutch national markers, 4,887 German national markers, and 1,018 Danish national markers.

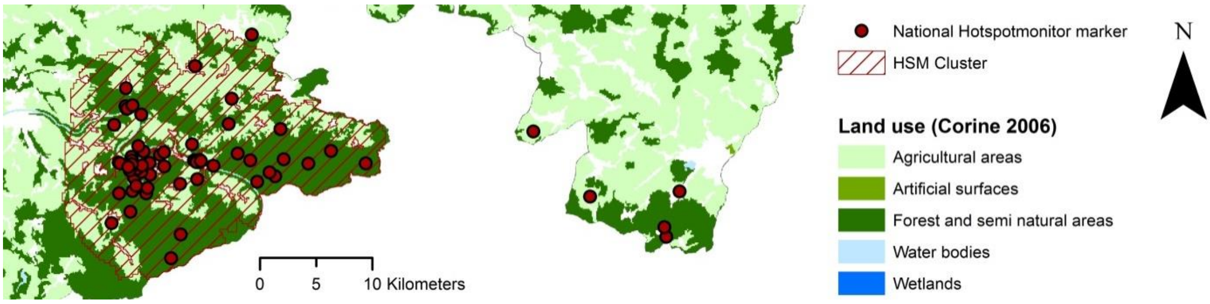
2.3 Defining clusters of collectively appreciated nature

- For a figure of the working dataset of clusters see Appendix III.
- A cluster is a natural area in which, or in the near vicinity of which, HSM markers are concentrated more dense than would be expected if these would be distributed across space evenly.

- Clusters are calculated per country separately; using only those national HSM markers located in the observed country and placed by respondents from that country – see Technical appendix part A for the data cleaning procedure. Calculation steps can be summarized as follows (for a more detailed description see Technical appendix part C).
    a. First, we calculate for each country the expected density of HSM markers per km² if these markers would be distributed across the country’s natural and agricultural land evenly – a hypothetical situation with no clustering. These values will be used to delineate clusters in step c.
    b. Then, per country, for each separate cell of a 250x250m raster grid we measure how many national HSM markers are located within a certain distance radius surrounding

the observed grid cell (3,073m for NL; 5,000m for DE; and 3,801 for DK[5]). The outcome is the density value of #HSM markers/ km² within the search radius surrounding a grid cell.

c. The next step is that we merge all adjacent grid cells with a value for #HSM markers/km ² within the search radius which exceeds the hypothetical value that would have been found if all markers would have been distributed across space. This merge defines our initial clusters.
d. For extra spatial precision of the clusters' fit on nature an overlay with Corine 2006 natural and agricultural land use data is made.
e. The clusters are quite consistent across the three countries analyzed.[6]
f. For each cluster the following figures are calculated:
    i. size in hectares
    ii. number of national Hotspotmonitor markers contained
    iii. HSI, which, as discussed later in this report in detail, essentially is a location quotient for the appreciation of a cluster's natural area.



View on one of the HSM clusters, which is located near the German border south-east of Dresden and along the Elbe. The figure shows that the markers on the right hand side of the figure are not concentrated densely enough to form a cluster; whereas the markers on the left are located in a cluster (which is shaded red). The clustering method is based on the concentration of HSM markers purely. However, an overlay with land use data is made to ensure a precise fit of the clusters on natural areas (including agricultural land).

[5] For the motivation for using different search radii see Technical appendix part C.

[6] The amount of markers is not proportional to the natural and agricultural land area of the three countries analyzed – consequences of this fact and how these are dealt with are discussed in high detail in the Technical appendix part C.

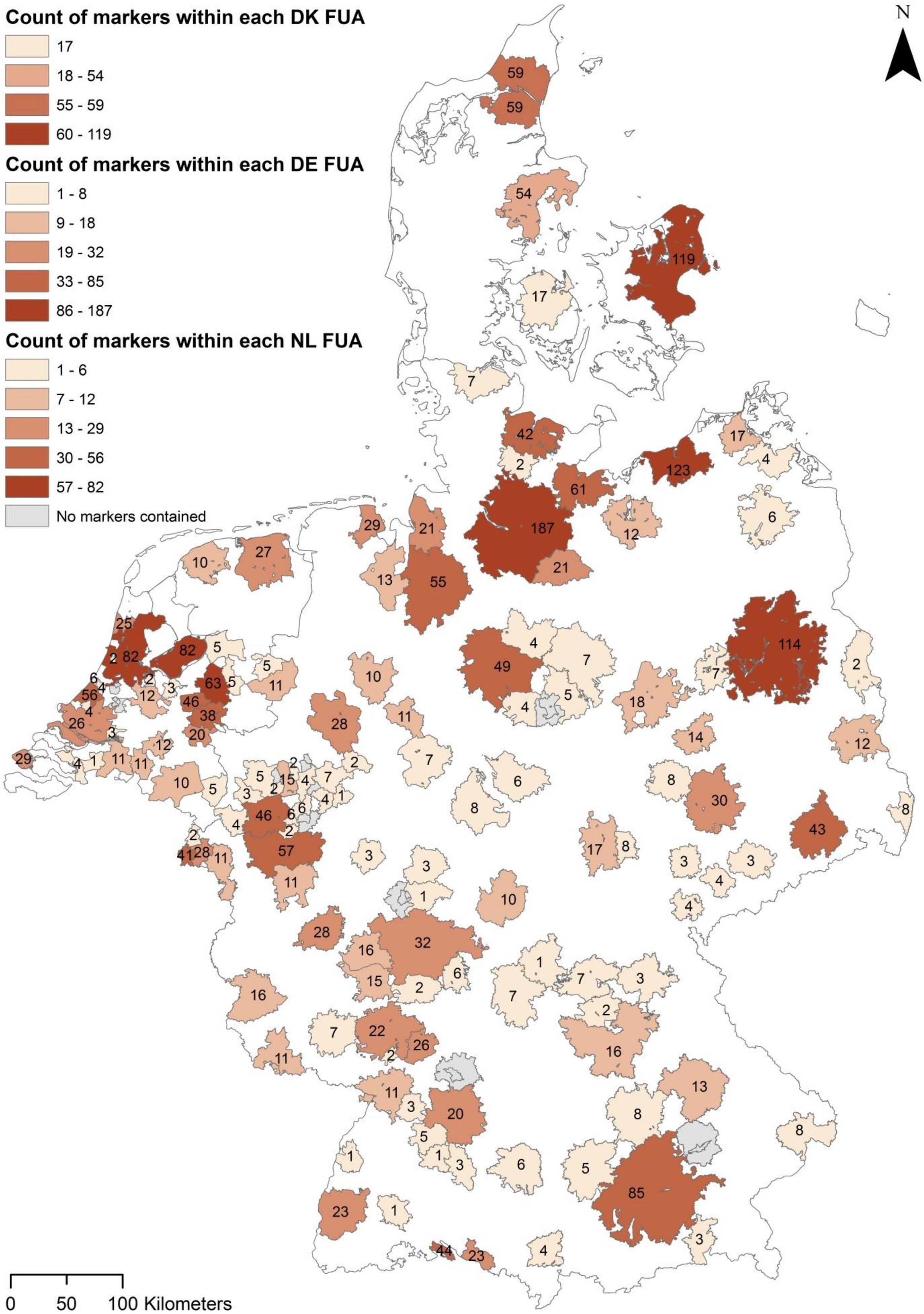
3 - Results

List of indicators

Indicator No.	Label
1A	Number of HSM markers within FUA
1B	Number of HSM markers within FUA+10km buffer surrounding it
2	HotspotIndex (HSI) FUA-level
3	FUA populations' mean distance to nearest HSM cluster
4	FUA populations' mean nearest HSM cluster's size/distance
5	FUA populations' mean nearest HSM cluster's (appreciation weighted size)/distance

- <u>Note on the calculation of indicators 3, 4, and 5</u>: their values are aggregates of values calculated at population grid cell level; these latter values influence on the outcome on FUA level corresponds with the number of inhabitants of each population grid cell.

Indicator 1A – Number of HSM markers within FUA

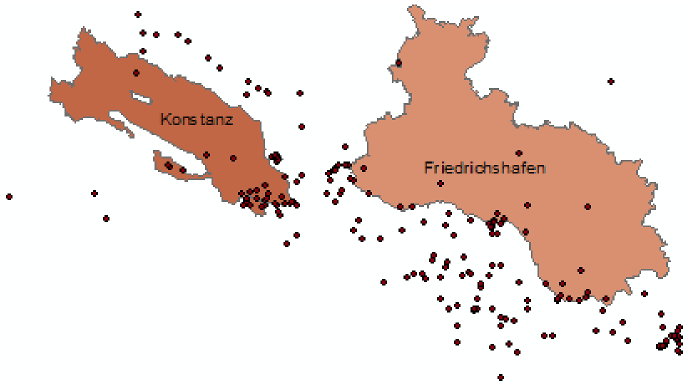


- **Rationale**: well-being may be positively related to the absolute intensity of collective appreciation for natural areas in a FUA.
- **Data**: national HSM markers
- **Calculation**: a count of markers contained by each FUA.

- **Pros**: easy interpretation
- **Cons**: as shown in the indicator-figure this indicator may undervalue amenity levels in FUAs as markers may be located just outside FUA (therefore use distance to clusters, or include markers in the count if within X kilometer of the FUA. Note that the numbers of markers per FUA are not comparable across countries because per country N markers differ. Also, larger areas have a larger potential for including many markers.

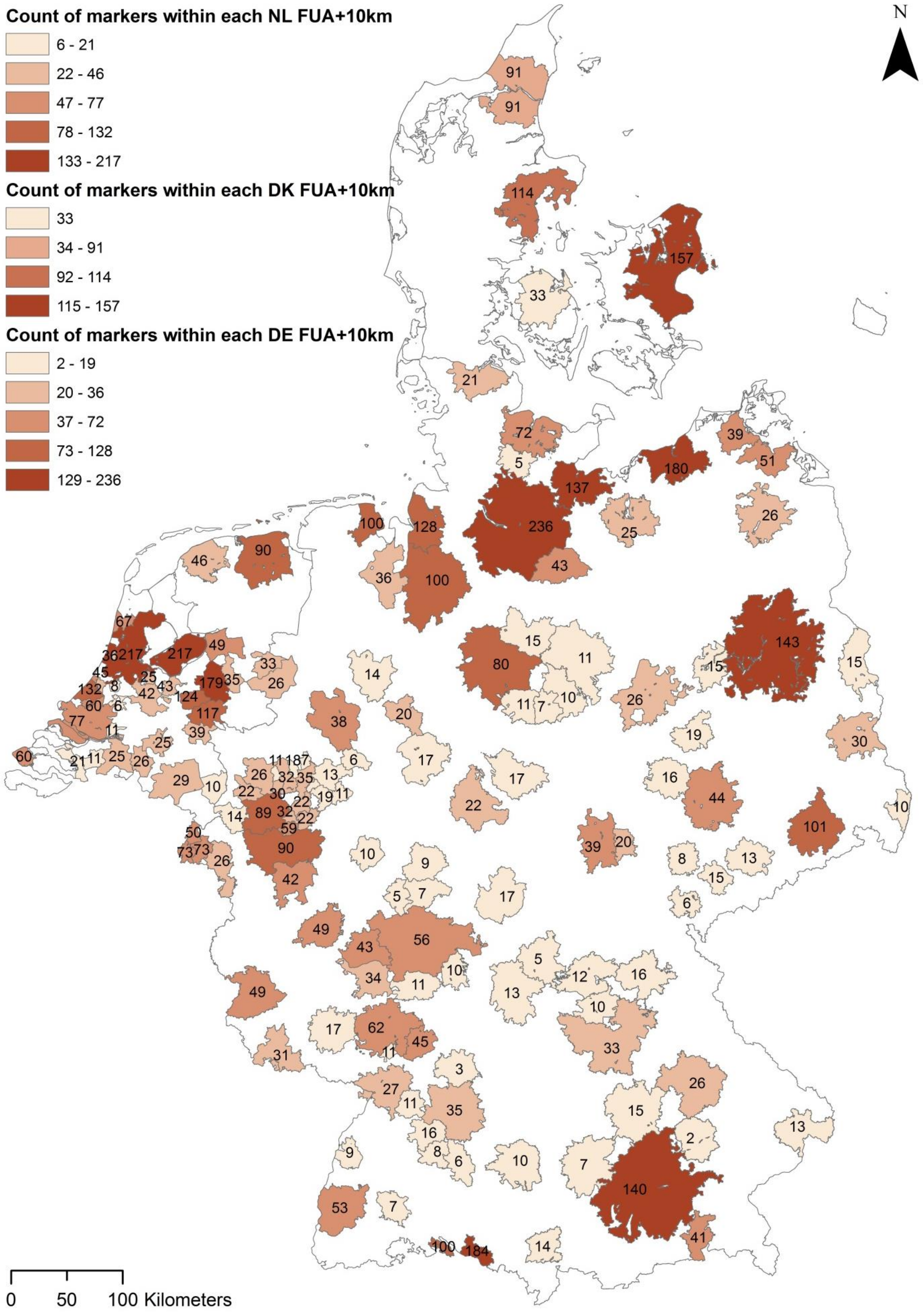
- **Alternative specifications of the indicator**: **(1)** the number of markers per FUA as a percentage of the total markers for the country in which the FUA is located. Using this specification the values reported for FUAs are better comparable across countries. One thing that hampers the interpretation of such indicator, however, is that for example in Germany appreciation is more diffusely spread over natural areas than in the Netherlands – partly due to the choice set of natural areas that German HSM respondents can choose from when placing a national marker on the map being larger than the choice set for Dutch HSM respondents as the Netherlands is smaller and contains less natural areas. **(2)** Indicators 1A could be scaled to reflect at FUA-level the number of markers per capita.

- **Issue in spatial measurement**: FUAs seem to typically exclude large natural areas. While these external natural areas may be appreciated intensively, ranking FUAs by HSM markers included within their borders may underestimate the natural amenity level associated with living in a FUA nearby an highly appreciated external natural area. Indicators 1B and 3/4/5 address this issue.



Two FUAs in Germany. Friedrichshaven is located *close* to more HSM markers than Konstanz is, indicating a higher natural amenity level– however, as the figure shows, a count of only markers *within* both FUAs does not reveal this.

Indicator 1B – Number of HSM markers within FUA+10km buffer surrounding it



- **Rationale**: FUAs seem to typically exclude large natural areas. While these external natural areas may be appreciated intensively, ranking FUAs by HSM markers included within their borders may underestimate the natural amenity level associated with living in a FUA nearby an highly appreciated external natural area. Also, FUAs borders are located on land, while a significant share of HSM markers is located on water right next to a beach – and as such coastal areas are somewhat undervalued when indicator 1A is used. For these two reasons the count of national markers within each FUA, as measured in indicator 1A, is extended to include also markers within 10km of the FUA. For several regions this indicators value differs from the value for indicator 1A.

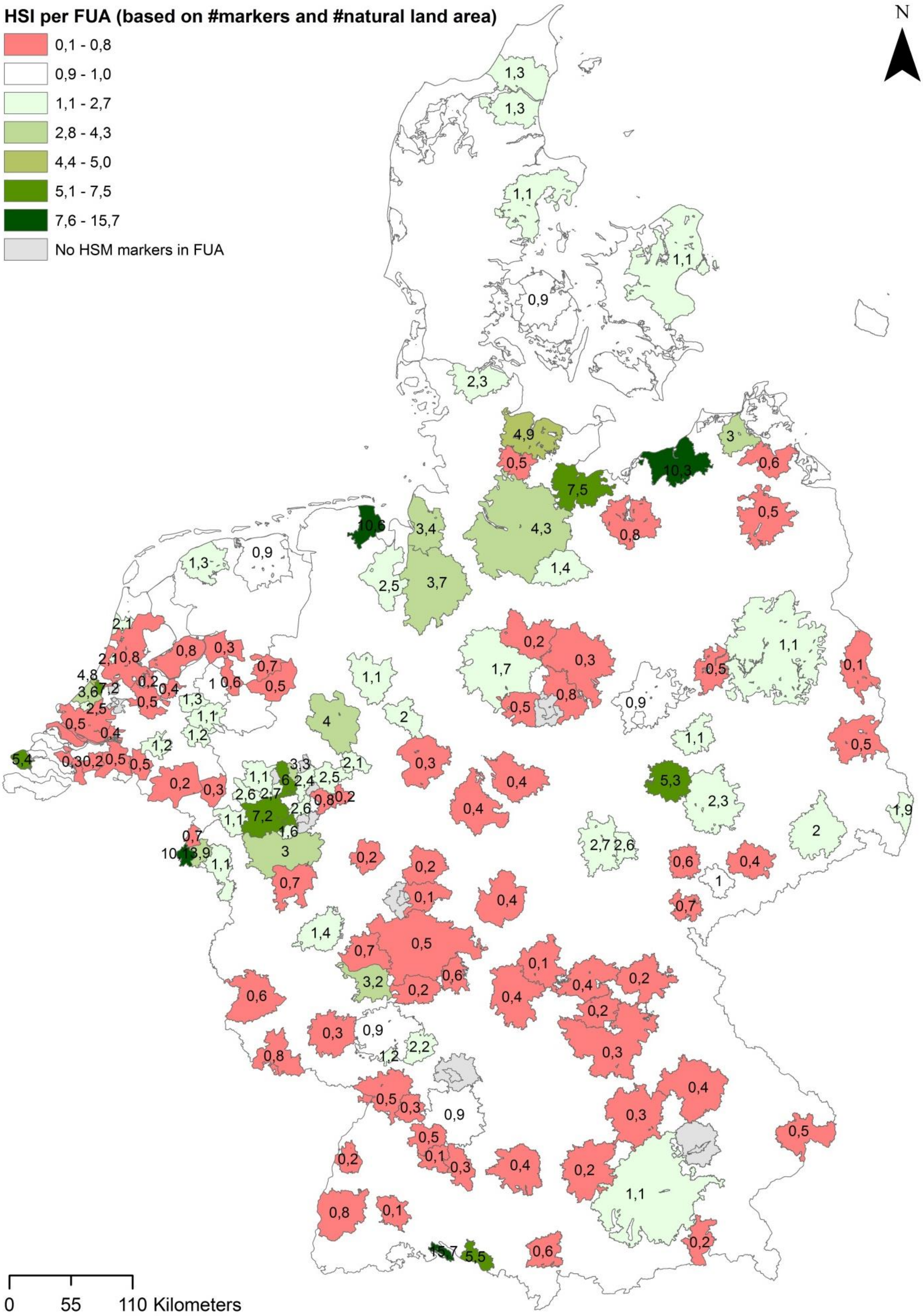
- **Data**: national HSM markers

- **Calculation**: a count of markers contained by each FUA.

- **Alternative specifications of the indicator (same as for Indicator 1A)**: **(1)** the number of markers per FUA as a percentage of the total markers for the country in which the FUA is located. Using this specification the values reported for FUAs are better comparable across countries. One thing that hampers the interpretation of such indicator, however, is that for example in Germany appreciation is more diffusely spread over natural areas than in the Netherlands – partly due to the choice set of natural areas that German HSM respondents can choose from when placing a national marker on the map being larger than the choice set for Dutch HSM respondents as the Netherlands is smaller and contains less natural areas. **(2)** Indicators 1A could be scaled to reflect at FUA-level the number of markers per capita.

- **<u>Issues unique for indicator 1B compared to indicator 1A</u>**:
- Buffering FUAs to count HSM markers somewhat violates the idea of a FUA as a coherent spatial unit. Also the buffer used here is of an arbitrary extent.
- If the percentage share in nationwide appreciated nature is calculated the per country sum of shares for the FUAs doesn’t add up to 100% anymore because of double counting.
- At the level of FUA inhabitants including HSM markers outside FUA borders in the count may be sensible as they may live near the border of a FUA. One option is to measure for each household within a FUA the amount of markers within a certain radius and aggregate the outcome to FUA-level.

Indicator 2 – HotspotIndex (HSI) at FUA-level



- **Rationale**: it seems meaningful to measure the ‘actual amenity value’ of each FUA’s nature in a standardized fashion. This can be done using the HSI, which reflects the ‘actual amenity value’ of a FUA’s stock of nature. A HSI > 1 indicates that the FUA’s nature is appreciated more intensively than would be expected given the FUA’s share in the countrywide stock of natural areas. For example, a FUA with a 1% share in the nationwide stock of natural areas would receive 1% of the countrywide HSM markers if appreciation for nature would be distributed equally across natural areas. However, if it would receive 2% of the markers, which is twice as much as expected given its country-wide shares in natural areas and an assumed non-clustered pattern of HSM markers, the HSI is 2.

- **Data**: National Hotspotmonitor markers; Corine 2006 land cover 100x100m grid

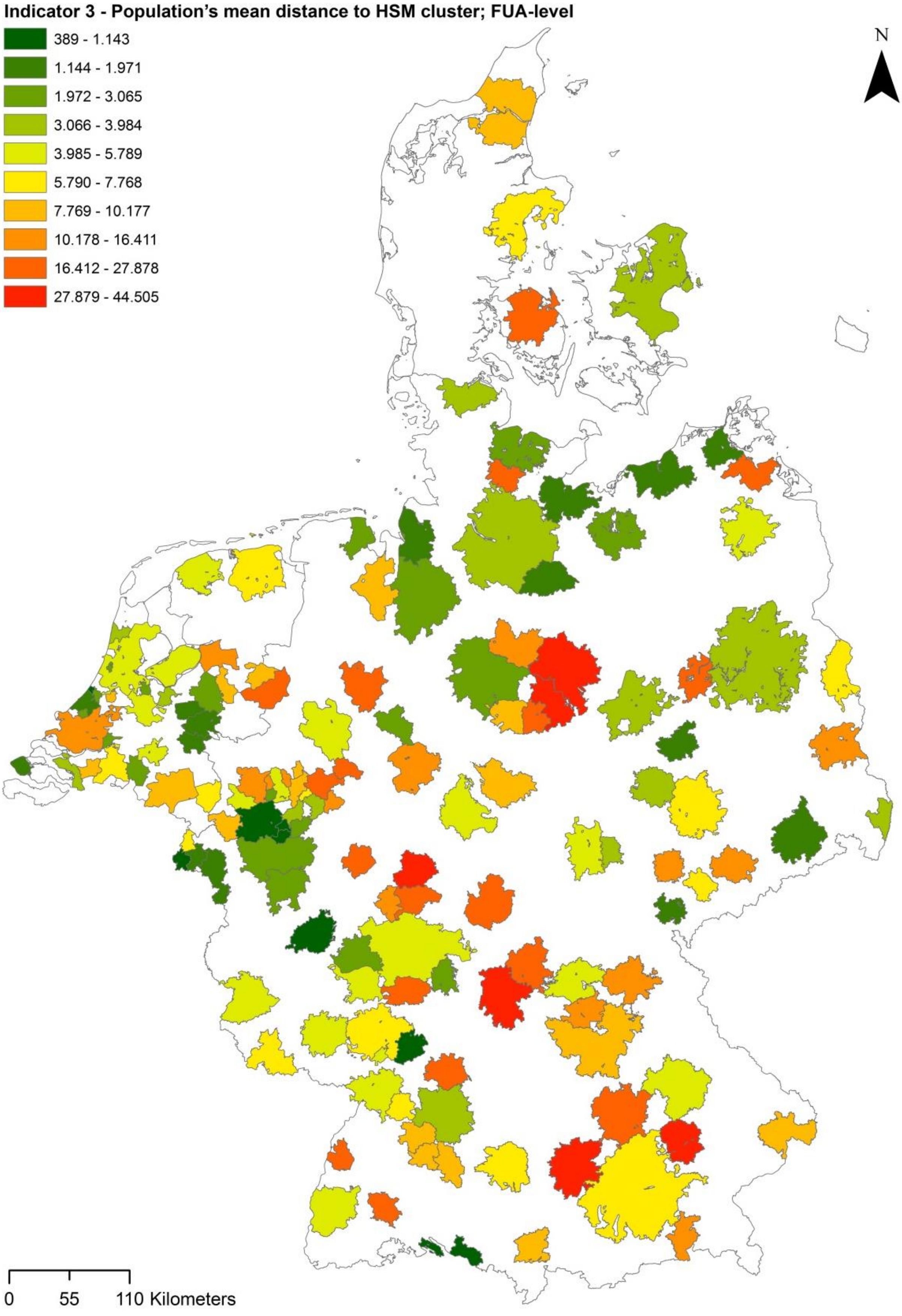
- **Calculation**:
  (#markers in FUA / (hectares natural land in FUA / hectares natural land country * #markers country))
  The left hand side of the equation gives the number of markers that would be expected in a FUA given its share in its country’s natural land area if appreciation (markers) would be spread equally across natural space.

- **Notes on calculation matters**
- <u>A location quotient style equation gives exactly the same outcome</u> as the HSI equation above: (N markers in FUA / FUA nature area) / (Countrywide N markers / Country nature area)
- See Technical appendix for more information on how natural land area per FUA is calculated using Corine 2006 land cover data.
- For more insight in the raw data on which the HSI values shown in the indicator-figure are calculated see the figure in Appendix I.

- **Pros**: standardized measurement of actual amenity value across FUAs from multiple countries, yielding comparable outcomes.
- **Cons**: large areas are disadvantaged if clusters are highly concentrated, and if concentration overlaps with small FUAs then these receive a relative high HSI-score.

- **Issue in spatial measurement**: similar to 1A: the appreciation for large natural areas external but nearby to a FUA is not included in this indicator as FUA borders are used.

Indicator 3 – FUA populations' mean distance to nearest HSM cluster



- **Rationale**: Euclidean distance to the nearest cluster may be negatively related to well-being.
- **Data**: HSM clusters (see Technical appendix C); OECD population statistics 1x1km grid
- **Calculation**: the distance to the nearest cluster is multiplied with the population in each grid cell, the sum of these values is calculated at FUA-level and then divided by each FUA's total population.

- Notes on calculation matters

- Instead of distance between a population grid cell to the nearest cluster the distance between a population grid cell's *centroid* and the nearest cluster is measured.
- A nearest cluster may be located in another country.
- Whether a nearest cluster is located in the same FUA as the population point is saved in a dummy variable. Using this dummy the figures in the table below are calculated.

- **Pros**: a nearest cluster may be located outside FUA borders, easy interpretation, spatially explicit as the location of a FUA's population at 1x1km grid centroids is accounted for
- **Cons**: measures distance to the nearest cluster only while an aggregated measure of distance to multiple clusters, weighted for distance decay, may provide additional information relevant to indicating well-being. Instead of Euclidean distance, travel time or network distance could be used.

Table 1 – Location of nearest cluster: inside or outside FUA

	NL	**DE**	**DK**
% of population whose nearest HSM cluster is located *outside* their FUA	12,7%	17,3%	0,9%
% of population whose nearest HSM cluster is located *inside* their FUA	87,3%	82,7%	99,1%
Total	*100,0%*	*100,0%*	*100,0%*

Table 2 - FUAs with the highest and lowest scores on Indicator 3, ranking per FUA class

Indicator 3 = a FUA's population's mean distance to the nearest HSM cluster

FUA class 1

Highest amenity level	*Score*	*Lowest amenity level*	*Score*
Maastricht (NL)	389	Salzgitter (DE)	25,416
Solingen (DE)	407	Neumünster (DE)	25,661
Konstanz (DE)	609	Brandenburg an der Havel (DE)	27,065
Friedrichshafen (DE)	764	Marburg (DE)	33,449
Katwijk (NL)	819	Landshut (DE)	37,957

FUA class 2

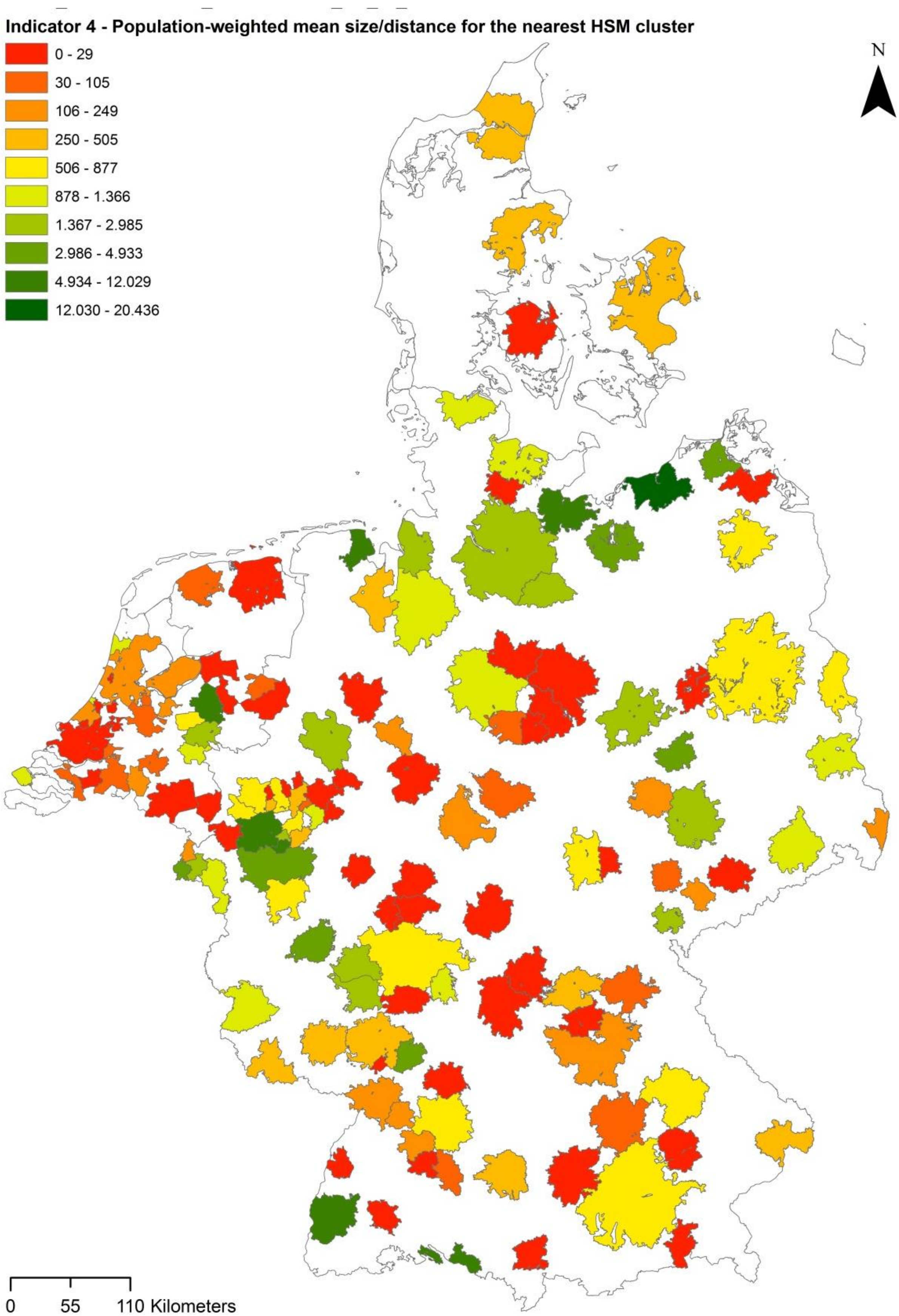
Highest amenity level	*Score*	*Lowest amenity level*	*Score*
Heidelberg (DE)	590	Fulda (DE)	25,054
Leverkusen (DE)	968	Giessen (DE)	27,878
Koblenz (DE)	1,143	Würzburg (DE)	35,180
Rostock (DE)	1,400	Braunschweig (DE)	36,986
Nijmegen (NL)	1,661	Wolfsburg (DE)	44,505

FUA class 3

Highest amenity level	*Score*	*Lowest amenity level*	*Score*
Düsseldorf (DE)	723	Eindhoven (NL)	9,431
Dresden (DE)	1,639	Duisburg (DE)	11,008
s-Gravenhage (NL)	1,880	Rotterdam (NL)	11,937
Aachen (DE)	1,926	Dortmund (DE)	18,221
Bremen (DE)	2,397	Augsburg (DE)	34,734

All FUAs in class 4	*Score (high amenity to low)*
Köln (DE)	2,577
Hamburg (DE)	3,744
Kobenhavn (DK)	3,796
Stuttgart (DE)	3,946
Berlin (DE)	3,984
Frankfurt am Main (DE)	5,227
Amsterdam (NL)	5,431
München (DE)	7,700

Indicator 4 – FUA populations’ mean nearest HSM cluster’s size/distance



- **Rationale**: Euclidean distance to the nearest cluster may be negatively related to well-being, an effect that may increase with the amount of natural area in a cluster.

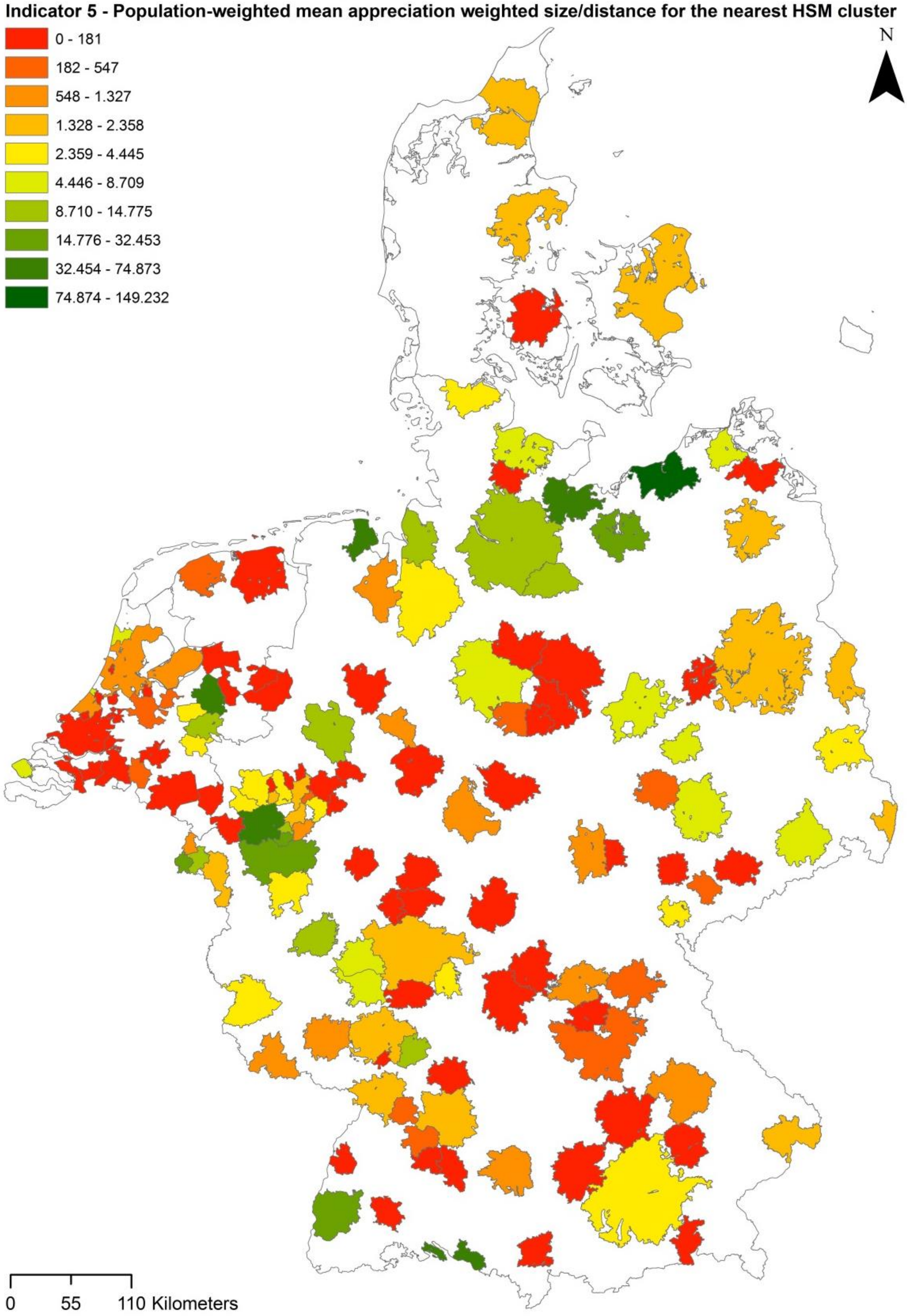
- **Data**: HSM clusters (see Technical appendix C); OECD population statistics 1x1km grid

- **Calculation**: the nearest cluster's size / distance is multiplied with the population in each grid cell, the sum of these values is calculated at FUA-level and then divided by each FUA's total population.

- Notes on calculation matters
- Instead of distance between a population grid cell to the nearest cluster the distance between a population grid cell's *centroid* and the nearest cluster is measured.
- A nearest cluster may be located in another country.
- Whether a nearest cluster is located in the same FUA as the population point is saved in a dummy variable.

- **Pros**: more comprehensive than indicator 4, spatially explicit as the location of a FUA's population at 1x1km grid centroids is accounted for
- **Cons**: (same as for indicators 3 and 5) measures distance to the nearest cluster only while an aggregated measure of distance to multiple clusters, weighted for distance decay, may provide additional information relevant to indicating well-being. Instead of Euclidean distance, travel time or network distance could be used. A country with relative large natural areas (Germany) effectively determines the scale of the map legend because the distance variable is included in the calculation of this indicator (analyzing regional HSM markers could prove useful in not undervaluing the amenity value of nature in FUAs near smaller natural areas).

Indicator 5 – FUA populations' mean nearest HSM cluster's (HSI * size)/distance



- **Rationale**: Euclidean distance to the nearest cluster may be negatively related to well-being, an effect that may be propelled by a higher intensity of appreciation for natural area in a cluster.

- **Data**: HSM clusters (see Technical appendix C); OECD population statistics 1x1km grid

- **Calculation**: the nearest cluster's (size * HSI) / distance is multiplied with the population in each grid cell, the sum of these values is calculated at FUA-level and then divided by each FUA's total population.

- Notes on calculation matters
- Variation: indicator * cluster within FUA dummy
- Nearest clusters from other countries are included in the calculation
- Population grid centroids used instead of grid cells
- Whether a nearest cluster is located in the same FUA as the population point is saved in a dummy variable.

- **Pros**: utilizes more information (intensity of appreciation of cluster's nature – HSI) on the perceived amenity value of natural areas than indicator 4 does (clusters only).
- **Cons**: (same as for indicators 3 and 4) measures distance to the nearest cluster only while an aggregated measure of distance to multiple clusters, weighted for distance decay, may provide additional information relevant to indicating well-being. Instead of Euclidean distance, travel time or network distance could be used. A country with relative large natural areas (Germany) effectively determines the scale of the map legend because the distance variable is included in the calculation of this indicator (analyzing regional HSM markers could prove useful in not undervaluing the amenity value of nature in FUAs near smaller natural areas).

4 – Discussion points

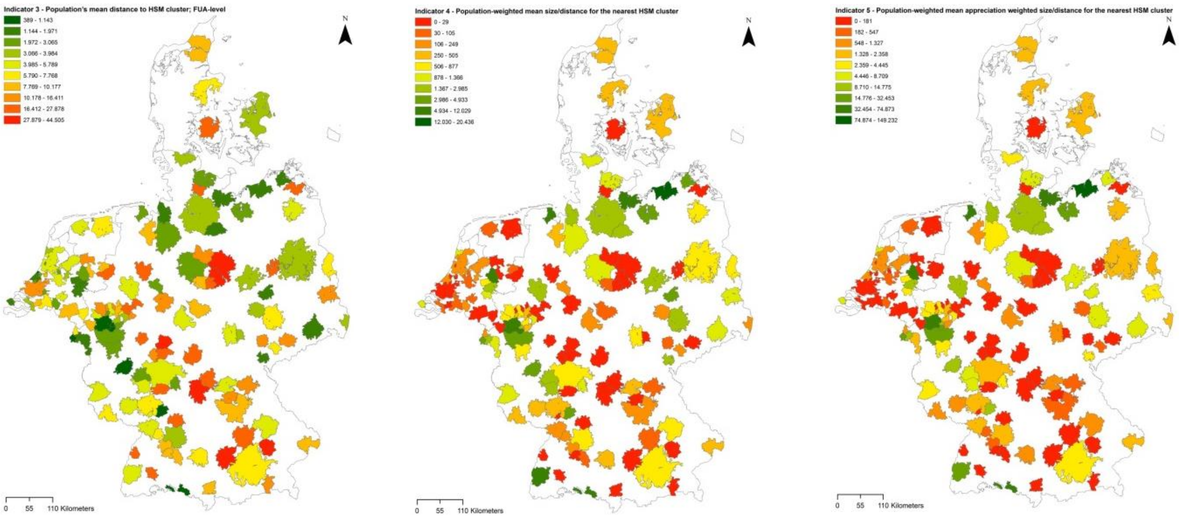
- The 1x1km population grid is rather coarse, spatially. In this report it is assumed that each grid cell its population is located at the cell's centroid. More fine measurement of the population's distance to the nearest HSM cluster can be achieved using a more fine grid.
- In this report only national clusters are used. Some FUAs may however contain few natural areas which are appreciated on a national scale, or may be located at a high distance from such nationally appreciated natural areas, while such FUAs may contain natural areas which are highly appreciated on a regional scale.
    a. When only national clusters' size (and HSI) are taken into account (indicators 4 and 5) Germany with its, compared to the Dutch and Danish natural areas, relative large natural areas effectively determines the scale of the map legend for the indicator. Analyzing regional HSM markers could prove useful in not undervaluing the amenity value of nature in FUAs near smaller natural areas.
- In this report we used FUAs with no distinction between core and hinterland as these were merged. When measuring urban well-being related to natural amenities it can however be relevant whether these natural amenities are located in the core or in the hinterland.
- In Indicators 3, 4, and 5 only the nearest cluster's value(s) is/are taken into account, while an aggregated measure of distance to multiple clusters, weighted for distance decay, may provide additional information relevant to indicating well-being.
    a. Problem: a HSI value can be constant due to the negative relation between number of markers included in a cluster and its area – and therefore an index which accounts for proximity to multiple clusters weighted for HSI would be difficult to interpret.
- Indicators 3, 4, and 5 were also calculated using distance squared instead of distance, however, due to outliers (clusters which were of a distance of e.g. 1 meter from population grid cell centroids), the output was the same as for the indicators measured with ordinary distance. For this reason these alternative specifications of the indicators mentioned were not included in the report. Nevertheless the use of including a distance decay function in indicator measurement seems worthy of further exploration.
- Creating inter-country consistent HSM clusters requires either a similar amount of national markers for the country observed relative to its natural and agricultural land area or scaling of the search radius within which potentially clustered markers are identified.
- This report focusses on FUAs quantity, or its population's mean proximity to, the collective's national HSM markers. An alternative would be to evaluate over which distance a HSM respondent in a FUA places a regional or local marker – taking a approach to measuring well-being which is more focused on individual subjective values.

Synthesis

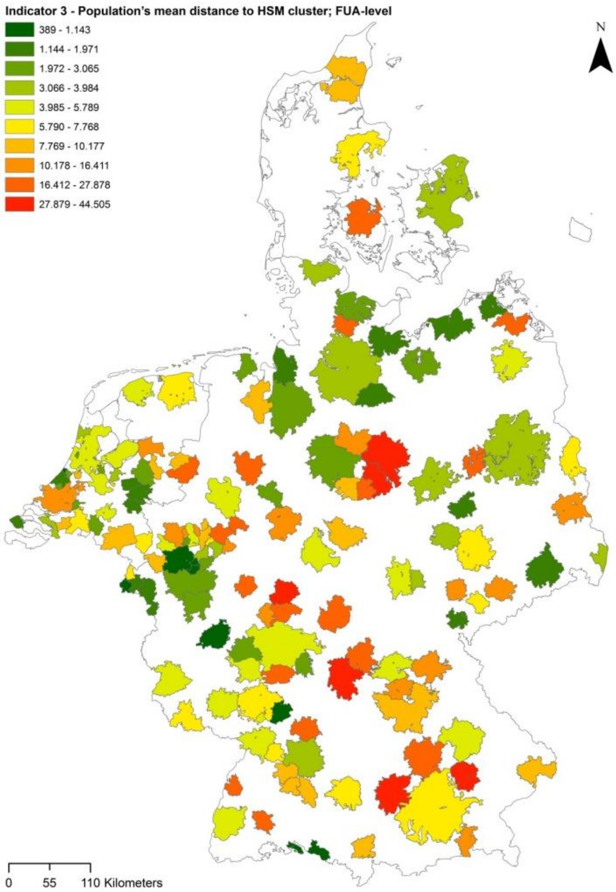
Here we give an overview of the maps of the different indicators discussed in the pages above, with the aim of coming to a preferred indicator.

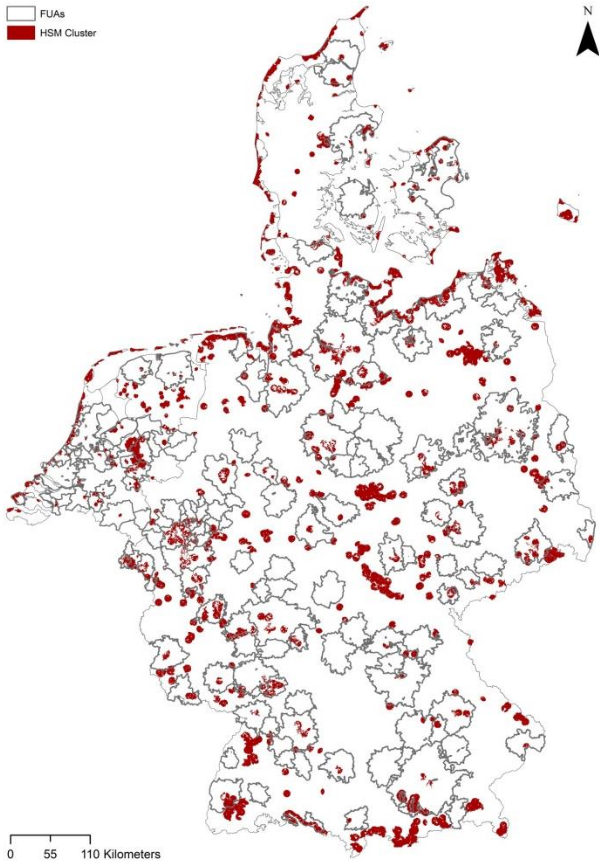
Overview of indicators: 1A, 1B and 2

Overview of indicators: 3, 4 and 5



With many indicators there is a trade-off: if the indicator is more comprehensive it may become more difficult to understand. On the other hand indicators that are too simple may give little information or even wrong or misleading information. Which of the six indicators shown above seems to best balance richness in information with ease of understanding? We feel that Indicator 1A and 1B suffer too much from the absolute scaling issues. Indicator 2, with the current scaling at the map, does not seem to give much information. However, the HotspotIndex seems to be a clear and useful indicator for differences in the 'actual amenity value' of the stock of natural areas in different FUA's. The differences between indicator 3,4 and 5 judged by the outcome maps seem to be quite modest, while 4 and 5 are relatively difficult to understand. Furthermore, especially for the Netherlands and Denmark indicator 3 seems to give more differentiation (i.e. may suffer less from scaling issues dominated by Germany). Indicator 3 is also the indicator that is easiest to understand: How near are people in an FUA to a national natural hotspot? If we for instance compare the resulting outcome map for Indicator 3 with the Appendix III map which shows the working set of national Hotspotmonitor clusters the score for every FUA is easy to understand / trace. It seems then that Indicator 3 balances the ease of understanding with richness of information best. Indicator 3 then seems to be a useful indicator for nature related well-being.





Indicator 3: distance to nearest HSM cluster figure (copy) **Appendix III map of HSM clusters (copy)**

Technical appendix

A - HSM marker data cleaning

1. Select from hsm_2013 all level 4 (national markers); N=7,656
2. Hand-select (lasso tool) all markers contained by NL / DE / DK and those within reasonable distance; export to 'hsm_2013_level4_workingset.shp'. 234 markers are removed
3. Split workingset into workingsets for NL/DE/DK separately because clusters have to be calculated for each country separately as these are national markers. Selection statement is 'lang=[nl/de/dk]'.
    a. N
        i. NL = 1,295
        ii. DE = 5,085
        iii. DK = 1,042
    b. Checked: subtotals sum up to the total number of markers.
4. Now for each country subset of markers remove all the national markers outside of the respective country's national borders and export the files to hsm_2013_level4_workingset[NL/DE/DK]_cleaned.shp'
    a. N removed observations
        i. NL = 31
        ii. DE = 198
        iii. DK = 24

B - Calculating the HotspotIndex at FUA-level

- **Data selection procedure:**
- Extract Corine 2006 100m raster by mask (FUAs)
- Raster to polygon
- Dissolve by gridcode (simplify polygons unchecked)
- Join labels
- Select 'natural land use' as everything except built and agricultural land use
- Dissolve all

Do this also for NL/DE/DK so the amount of nature in country borders can be calculated. Note that this excludes marine water bodies while these could be marked in the HSM. Therefore the HSI figures based on these data will be slightly biased.

- Clip FUAs by nature_selection_dissolve so only the FUA area covered by nature remains
- Then calculate hectares nature per FUA and join this back to the original FUA file

Do this also for NL/DE/DK to calculate the national part of the HSI hit chance.

C - Definition of HSM clusters

5. Calculate the km² area of each country's land surface, and # markers
   a. NL # hectares = 34,975 # markers = 1,264
   b. DE # hectares = 357,791 # markers = 4,887
   c. DK # hectares = 43,075 # markers = 1,018
6. Calculate cut-off points (#markers / #hectares country land)
   a. NL = 1,264 / 34,975 = 0.0361
   b. DE = 4,887 / 357,791 = 0.0137
   c. DK = 1,018 / 43,075 = 0.0236

- Create point density grids for each country's set of markers separately
  - Search radius for markers is:
    - NL: 3,073m
    - DE: 5,000m
    - DK: 3,801m
  - The NL and DK search radii are scaled to deviate from the DE search radius for three reasons. First, the #markers/hectare differs per country, leading to a different expected #markers to be found when calculating clustering based on markers within a certain radius. Second, adjusting the number of markers is not possible on in the short term, and adjustment through only selecting a higher cut-off point leads to cross-national differences in the sensitivity for the absolute number of markers in an area when clustering (min. markers could be 5 for a cluster in NL and 2 in DE – this degrades cross-national comparability of clusters). Third, because of the second reason another parameter should be adjusted for two of the countries to match with one country where the fit of clusters on markers is good. Visual inspection indicated that the fit of clusters on markers is good in DE. The benefit of using these different search radii for each country, and adjusting them to the DE fit, is that given the N markers for each country and the area associated with each of the search radii the # expected markers within a search radius is constant across the three countries. A disadvantage of this approach is that cluster-size is positively related with the extent of the search radius – as clusters' outer borders are at the distance of the search radius from the nearest point it is generated from. Thus DE will have larger cluster areas than NL when the clustering circumstances are the same. DE natural areas are, however, also of a larger scale than NL natural areas.
- For each country clusters are created from the point density grids using as a cut-off value the expected markers/km², which are calculated as:
  - # markers for country / # hectares of land in country = expected # markers / km²

The value for this variable gives the number of markers that would be expected on one km² if all markers would be distributed over the country's space evenly. Choosing this value as a cut-off value assumes that clustering of HSM markers is present in wherever area where more markers are concentrated than would be expected if markers were distributed across the country's space equally.

- Raster grid values containing the #markers within the search radius are multiplied with 10,000 because decimals will be lost after converting the raster grid to polygon.
- Extract raster cells by mask – the respective country's land area borders.
- Raster grid cells with a value >= cut-off value*10,000 are selected and converted to polygons (with simplify polygons unchecked).
- To each polygon containing clusters a double formatted field 'pdens' is added; it is calculated to contain the original #markers per km² within the search radius through dividing gridcode by 10,000.
- Merge NL/DE/DK clusters
- Dissolve, singlepart to multipart
- Overlay with Corine 2006 natural + agricultural area (thus only excluding built land use)
- Dissolve, singlepart to multipart
- Recalculate clusters areas in hectares
- Count markers in clusters
- Keep clusters with #markers > 1 (since 1 marker can still be random)
- Add country dummy to clusters
  - 9 FUA's are located in two countries (8/9 for a very large part in the second country). Only 1/9 is problematic to assign manually to one country: Cluster_ID = 4102. Two clusters, one in DK and one in DE have grown together. It is manually assigned to Germany. Its HSI will therefore be somewhat biased because part of it depends on DK (expected and actual) markers and part on DE (expected and actual) markers.
- Calculate clusters HSI.
  - Count national markers (see earlier figures)
  - Calculate hectares *natural and agricultural land* per country
    - NL natural and agricultural area = 3,048,329.9 hectares
    - DK natural and agricultural area = 4,050,424.8 hectares
    - DE natural and agricultural area = 32,935,064.3 hectares
      - (=NL+DE+DK-NL-DK); NL+DE+DK= 40,033,819 hectares
  - Calculate HSI, as:
    ([Cluster_markers]/([Cluster_HA]/ [Country_natagri_HA]* [Country_markers]))
- Keep clusters with HSI>=1

    - No clusters removed; 407 out of 407 clusters satisfy the selection statement
    - These are the final clusters
- Add FUA dummy to clusters
    - External file: union of FUAs (dissolved by name) and final clusters
    - Export to .xlsx
    - Create a concatenation of Cluster_ID and NAME (which contains the FUA name); [ClusterID]_[NAME]
        - Note that a single cluster may overlap spatially with multiple FUAs

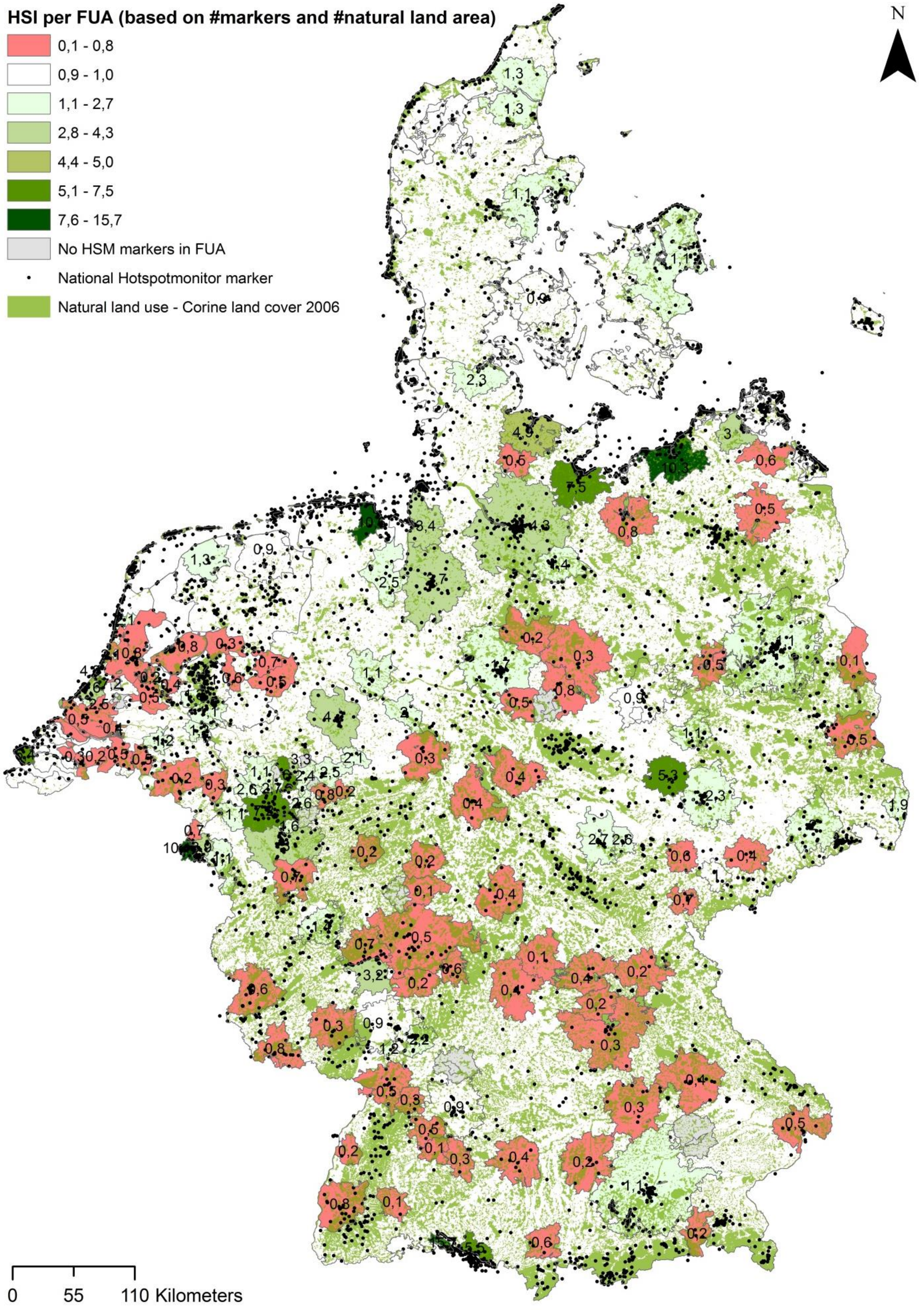
D - Population distance to clusters

- Population in 2006 on a 1x1km grid (*the centroids are used*); polygon shapefile
- Select all grid cells which intersect with a FUA (N = 115,142)
- Calculate XY centroids in ETRS89_LAEA coordinate system for each population grid cell
- Export to table to database file, and from this database file plot the XY as points and export to shapefile.
    - From now on it is assumed that these centroids are the actual locations of the population included in the centroids' grid cells; as such these will be referred to as population points
- Spatial join of population points to add the NAME of the FUA which it is located in
- Keep only points located within borders of a FUA
    - 7,756 points are removed as these are located outside of a FUA – as their grid cells are located partly inside and partly outside a FUA but with the centroid of the cell on the outside
    - The working set includes 107,386 population points, with a total population of 67,573,404
- Spatial join of population points to calculate distance to nearest cluster, and to add the cluster's values (the HSI and Cluster_HA are critical) to the population point
- Export to .xlsx
- Evaluate whether the nearest cluster is located in the same FUA as the population point
    - Create the same concatenation in the file with population grid cell of Cluster_ID (of the nearest cluster) and NAME (of the FUA within which the population point is located); [ClusterID]_[NAME]
    - Create a dummy: D_cluster_in_FUA and set all values to 0
    - Use vLookup to see whether there is a match; if so give the dummy value 1
        - This means that the nearest cluster is at least partly located in the same FUA as the population point
- Calculate indicators 3A and 3B.

  - For the variable distance (Euclidean; measured in meters) all 0's are replaced with 1's. A 0 is found when a population point is located in a cluster. By replacing 0's with 1's indicators based on equations which include cluster size or HSI divided by distance can be calculated.
  - Subtotals based METRO NR -this can be joined back to GIS without errors whereas a join based on NAME may lead to mismatches because some names spellings change when exporting/importing them using ArcGIS. METRO NR remains consistent.
- Join results back to shapefile based on FUA NAME based on METRO NR.

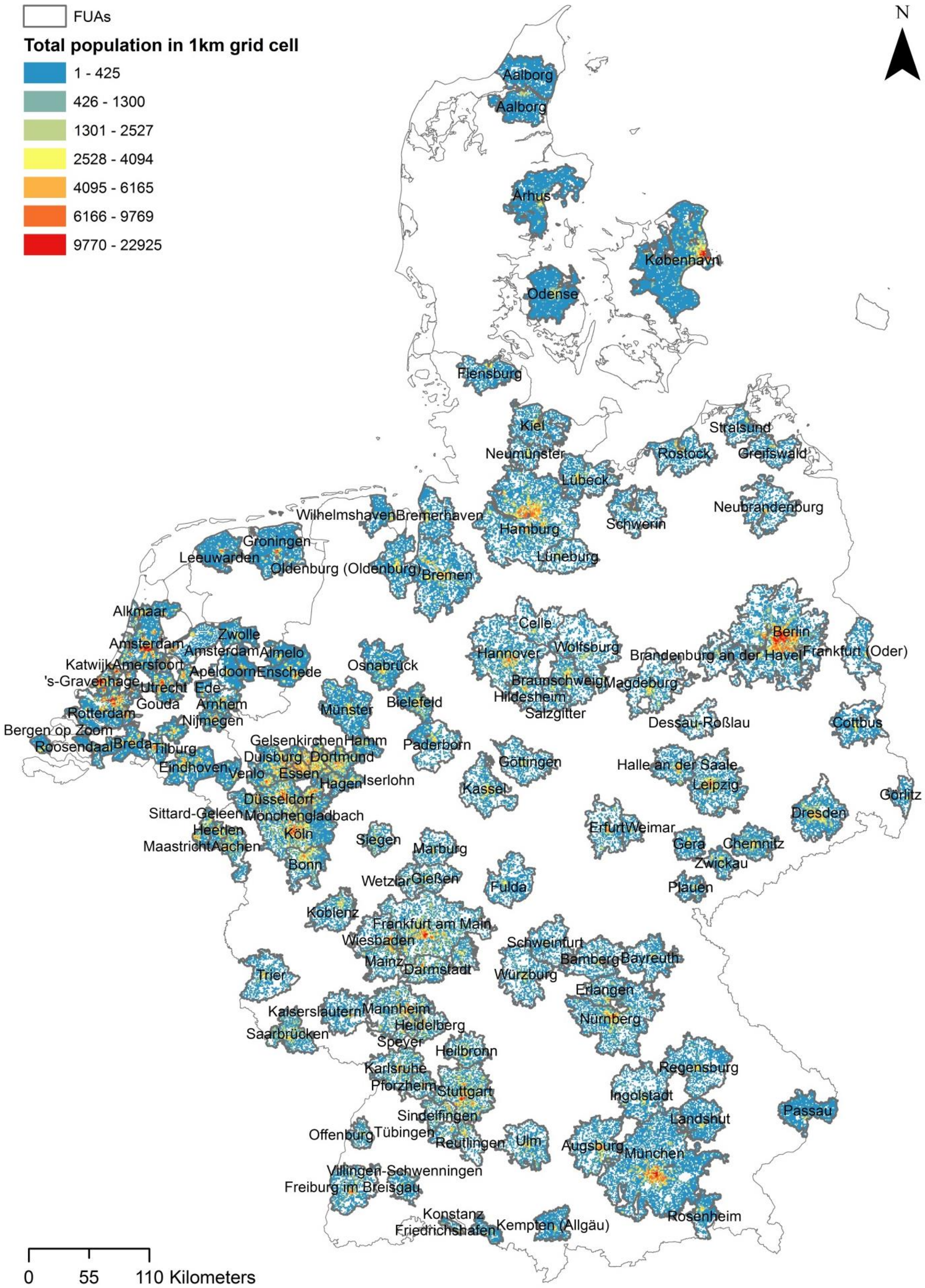
Appendix I

HSI values per FUA + data on the national HSM markers and natural land use



Appendix II

Population per 1km grid cell within FUA



Appendix III

Working dataset of HSM clusters